A new approach to using genetic data to define management units for Gulf of Mexico bottlenose dolphins

Karen K. Martien, Anna B. Sellas, Patricia E. Rosel, Barbara Taylor, and Randall S. Wells

ABSTRACT

Defining stocks for bottlenose dolphins in the Gulf of Mexico is complicated by the apparently continuous distribution of dolphins in these waters. While genetic studies promise rapid answers, most analytical methods are hindered by the need to begin with hypothesized strata. Long-term studies in the vicinity of Sarasota Bay, Florida allow hypothesized population structure to be based on extensive habitat and behavioral data, but such data are not available for most of the Gulf. We present a new analytical method, called *Boundary Rank* (*BR*), for generating population structure hypotheses in bottlenose dolphins and other continuously distributed species. We applied *BR* to an existing genetic dataset for bottlenose dolphins from Sarasota Bay and surrounding waters as an empirical test of the performance of this approach. We performed two analyses, one of which incorporated data from the long-term observational studies, the other of which did not. The results of both analyses were consistent with the results of long-term observational and photo-identification studies of population structure in the study area, suggesting that this method can be successfully applied in areas of the Gulf for which long-term observational data are not available. BR outperformed another available clustering method called *SAMOVA*. Estimated dispersal rates between the four major water bodies suggest that these areas represent demographically independent populations that warrant separate management under the MMPA, supporting previous behavioral and genetic analyses.

INTRODUCTION

The Gulf of Mexico is home to a number of photographic identification studies of bottlenose dolphins, including the longest-term study of wild bottlenose dolphin movements in

the world (Scott et al. 1990, Wells 1991, 2003, Wells and Scott 1999, 2002, Reynolds et al. 2000). Along the central west coast of Florida, findings of long term residency of dolphins within bays and sounds, coupled with repeated patterns of social associations, led to the description of geographically-based, multi-generational "communities" of dolphins, each community inhabiting home ranges with borders slightly over-lapping those of the adjacent communities (Wells 1986, 1994; Urian 2002). However, the understanding of population structure throughout the rest of the Gulf remains problematic. Nearly continuous distributions of bottlenose dolphins complicate differentiation into geographically-based stocks. NOAA Fisheries (NMFS) currently recognizes more than 30 putative stocks of bottlenose dolphins along the coast of the Gulf of Mexico (Waring et al. 1999), but most were established in the absence of appropriate empirical biological data, on the basis of geographical features alone. Throughout this paper we define a 'stock' to be a demographically independent population, in accord with the definition used by NOAA Fisheries under the Marine Mammal Protection Act (Wade and Angliss, 1997; Taylor, 1997).

Because very detailed long-term studies have not been conducted throughout the Gulf, there is considerable interest in using genetic methods to define stocks for bottlenose dolphins. However, most genetic analytical methods require the *a priori* definition of hypothesized management units. This initial stratification can strongly influence results (Martien & Taylor, in press). Here we present an analytical method, called *Boundary Rank* (*BR*), to generate hypothesized stocks using the genetic data. Our method uses hierarchical clustering to cluster samples based on their genetic similarity, resulting in a set of nested stock structure hypotheses. To evaluate the feasibility of this approach for defining population stocks for bottlenose dolphins in the Gulf of Mexico, we applied it to an existing genetic dataset from Sarasota Bay and

surrounding waters (Sellas et al, in prep.). This area has been extensively studied through long-term observational, photo-identification, and genetic work (Duffield and Wells, 1986, 1991, 2002; Scott et al., 1990; Wells, 1986, 1991, 2003; Sellas, 2002; Sellas *et al*., in prep.). The movement patterns and population structure of dolphins in this area are well understood, making it an ideal study system for an empirical performance test of new analytical approaches.

BR requires that genetic samples be grouped *a priori* into geographically small initial units, which serve as the smallest units used in the clustering. However, when samples are continuously distributed, as is often the case for marine mammals, initial units must be defined subjectively and therefore may impact the performance of the method. Expert knowledge of the behavior and fine-scale habitat preferences of the animals being studied can be very useful when defining initial units. However, such knowledge is not available in most areas of the Gulf of Mexico, nor is it usually available for most species of conservation concern. One of the goals of this study was to determine the importance of expert knowledge when defining initial units for a *BR* analysis of bottlenose dolphins in the Gulf of Mexico. Therefore, we performed and compared two separate analyses, one of which incorporated information gathered during the long-term observational and photo-identification studies into the definition of initial units, and the other of which did not.

Because BR results in a set of hypotheses, the additional step of estimating which putative stocks are demographically independent still needs to be taken. We accomplish this by estimating dispersal rates using the program Migrate (Beerli & Felsenstein 2001). We also compare the performance of BR to another clustering method available for continuously distributed species called SAMOVA (Dupanloup et al. 2002).

In addition to using the bottlenose dolphin genetic dataset as an empirical test of the performance of BR, we also performed extensive simulation performance testing of the method. We used the simulation model described in Taylor *et al*. (2000) to generate datasets for which we knew the true population structure. We then analyzed the simulated datasets using BR and compared the population structure hypotheses suggested by the method to the 'true' structure. The performance analyses considered a number of different scenarios known to commonly occur with real data sets, including unequal population sizes, unequal dispersal rates and uneven sampling effort. Details of the simulation performance tests can be found in the Supplement Information.

METHODS

Boundary Rank

Boundary Rank (*BR*) is an agglomerative clustering analysis that coalesces units on the basis of their genetic similarity. At each step in the clustering, we seek to merge the two units that are the most genetically similar and are therefore most likely to have come from the same biological population. We considered several different statistics to measure genetic similarity between units. We compared the performance (see Supplemental Information) using F_{ST} (Wright, 1932), f_{ST} (Excoffier *et al*. 1992) and χ^2 per degree of freedom (χ^2/dof). All measure the degree of genetic differentiation between two samples. ϕ_{ST} takes into account the evolutionary relationship between haplotypes as well as their frequencies, while the other two only consider haplotype frequencies. Performance was highest using χ^2/dof as the measure of genetic differentiation (See Supplemental Information), so that is the measure we used for the remainder of our analyses.

In order to make comparisons based on haplotype frequencies, the user must first group samples into geographically small initial units. Consequently, *BR* does require some *a priori* stratification of samples. What distinguishes these initial units from the strata used in a traditional hypothesis-testing analysis is their geographic scale; the initial units used in a *BR* analysis will be smaller than the strata typically defined for a traditional hypothesis test. With traditional hypothesis tests of population structure, the goal is to stratify the samples in a way that mimics the actual population structure as closely as possible. The goal with *BR*, on the other hand, is to define initial units that can be combined into hypothesized populations by *BR* in the way that is most consistent with the genetic data. The initial units should be as geographically small as is feasible to minimize the risk of defining an initial unit that spans a population boundary. However, if the initial units are too small then they will not contain enough samples to allow for meaningful frequency-based comparisons. In many cases, defining initial units will be easily accomplished by simply combining all samples taken from the same sampling site. However, when samples are not clustered geographically, the definition of initial units will be subjective and sensitivity analyses should be used to ensure that the results of the analysis are not heavily dependent on the choice of initial units.

The main feature that distinguishes *BR* from other genetic clustering methods is that it allows the researcher to place geographic constraints on the clustering so that the clusters produced by the analysis are geographically contiguous. Such a constraint is particularly valuable in ecological and applied studies where the goal is to investigate fine-scale population structure. In such studies, the population units of interest typically exchange dispersers at a high enough rate to eliminate any phylogeographic signal, rendering unconstrained clustering methods ineffective. Nonetheless, the dispersal rate is often still low enough that the population

units are demographically independent. Similar geographically constrained clustering methods

have been used in the past in other applied studies aimed at identifying management units. For

instance, York et al. (1996) employed a geographically constrained clustering approach in their

efforts to use trend data to identify management units for Steller sea lions.

The program uses a connectivity matrix to indicate which sampling sites can be directly

clustered with one another. A one ('1') is used in the connectivity matrix to indicate that two

sites can be clustered and a zero ('0') to indicate that they cannot. The fact that two sampling

sites are not considered 'connected' in the connectivity matrix does not imply that an individual

could not move between those two sampling sites. Rather, it simply indicates that those two

sampling sites could not be placed in the same conservation unit without including other

intervening sampling sites. By allowing incorporation of information on the behavior and

movement patterns of the species into the analysis, the connectivity matrix constrains the

analysis so that the population structure hypotheses suggested by BR are plausible given what is

already known about the biology of the species. or non-migratory animals such as many coastal

bottlenose dolphins, connectivity will likely be limited to geographically adjacent sites. For

migratory species, the researcher may wish to allow connections between non-adjacent sites.

Such a connectivity matrix would allow for the possibility of, for instance, the intrusion of one

breeding population into another along a migratory corridor or on the feeding ground.

At the beginning of the analysis, each sampling site represents a hypothesized

conservation unit. The genetic differentiation (χ^2/dof) is calculated between all connected pairs

of hypothesized units, and the most genetically similar pair of connected units (i.e., those

exhibiting the least genetic differentiation) is identified. The boundary separating those two

putative units is the lowest ranking boundary. It is removed and the two putative units are

coalesced, thereby reducing the number of units by one. The connectivity matrix is updated to reflect the geographic relationship of the new conglomerate unit and the genetic differentiation matrix is recalculated. The program outputs the identity of the two units that have just been combined, as well as the genetic differentiation between them. The process is then repeated, always removing the lowest ranking remaining putative boundary, until only a single boundary remains that divides the samples into two hypothesized units. The result is a nested set of plausible population structures, each containing one fewer units than the previous. This becomes the researcher's set of hypotheses concerning population structure. The method was extensively performance tested using simulations and found to perform well at suggesting population structure hypotheses that are consistent with the actual population structure of the samples (Supplemental Information).

Application to Gulf of Mexico Bottlenose dolphins

The data

We applied *BR* to the genetic dataset presented in Sellas et al. (in prep.). The dataset consisted of 450 bp of mitochondrial control region sequence from animals from Charlotte Harbor (n = 51), Sarasota Bay (n = 54), Tampa Bay (n = 46) and coastal waters of the Gulf of Mexico off west central Florida up to 9.3 km from shore (n = 56) (Figure 1). Samples were collected by projectile biopsy darts or during surgical biopsy for health assessment (in Sarasota Bay, only; Sellas *et al*, in prep.; Wells et al. in press) and included many animals with sighting histories. Sellas et al. (in prep.) found little genetic mixing between areas stratified using full behavioral data, which agrees with the long-term residency suggested by previous photographic identification studies in these areas – Sarasota Bay (Scott et al. 1990, Wells 1986, 1991, 2003), Charlotte Harbor (Wells et al. 1996a) and Pine Island Sound (Wells et al. 1997) combined,

Tampa Bay (Wells et al. 1996b), and the adjacent Gulf of Mexico waters (Fazioli and Wells

1999), and previous genetic data (Duffield and Wells, 1986, 1991, 2002).

Defining initial units

      We examine the impact of the definition of initial units by performing two separate

analyses, the 'naïve' and 'expert' analyses. For the naïve analysis, initial units (Figure 2a) were

defined by a researcher (KKM) who had a general knowledge of dolphin behavior, but no

specific knowledge of the individual sighting histories of the dolphins included in this study, nor

of the fine-scale habitat differences across the study area. The researcher did take into account

gross habitat differences (i.e., inland versus coastal waters) and geographic features (e.g.,

peninsulas and islands that divide inland waters into discrete bays). Initial units grouped samples

collected in the same general habitat type (inland or coastal water) that were clustered together

geographically (judged subjectively by eye; Figure 2a).

      Initial units used in the expert analysis (Figure 2b) were defined by an author (RSW) with

extensive knowledge of the behavioral patterns of bottlenose dolphins in the study area, the

microhabitats of Tampa Bay, Sarasota Bay and Charlotte Harbor, and the individual sighting

histories of many of the individuals used in the study (Wells 1991, 2003; Wells et al. 1987).

This background knowledge defined initial units more consistent with the behavior of the

animals. For instance, in the naïve analysis, animals that were known to live primarily in

Sarasota Bay were placed in the same initial unit (Unit 7 in Figure 2a) as some animals that have

been sighted primarily in Tampa Bay. For the expert analysis, the known Sarasota and Tampa

Bay animals were always placed in separate initial units. The initial units within Tampa Bay

were defined primarily on the basis of habitat differences for the expert analysis; the animals

from Boca Ciega Bay (Unit 1 in Figure 2b), which is a very shallow, complex area of small

islands and channels, were placed in a separate unit from those sampled in the deeper, open waters around the mouth of Tampa Bay (Unit 2 in Figure 2b). Similarly, the animals from the shallow habitat around the mouth of the Manatee River (Unit 3 in Figure 2b) were also placed in a separate unit from the deep-water samples.

A genetic analysis of population structure may be misled by the inclusion of closely related animals in the dataset, particularly if the probability of sampling relatives differs across the study area. *BR* may be particular vulnerable to the biases introduced by related individuals because the initial units are geographically small and, consequently, will often be comprised of relatively few samples. The average relatedness for the bottlenose dolphin dataset did not differ significantly between the four populations (Sellas *et al*., in prep.). However, the dataset did include several animals from Sarasota Bay that were known from long-term observational data to be mother-calf pairs. The calves of these pairs were excluded from the dataset for the expert analysis but not for the naïve analysis, since the latter analysis reflects the results that could be obtained in any field study where long-term data were not available to define first order relatives. Including these mother-calf pairs in the naïve analysis allowed us to test for possible biases that could result from their inclusion.

Connectivity matrices

The connectivity matrix for both the naïve and expert analyses was drawn to reflect a pure stepping stone model; in other words, connections were drawn only between initial units that are geographically adjacent (Figures 2a and b). This resulted in the definition of geographically contiguous units. There was one exception to the assumption of a stepping stone model: in both analyses, a connection was allowed between the southern end of Sarasota Bay and the northern end of Charlotte Harbor. This connection allows for the possibility that inshore,

estuarine populations are more closely related to each other than to intervening coastal populations. Such a pattern could be the result of habitat selection by the dolphins and is plausible given what is known about population structure of other species and in other regions (Wells and Scott 1999). In addition, there is a waterway connection between Sarasota Bay and Charlotte Harbor that would allow animals to move between these two areas without passing through the coastal population, and dolphins have been documented making this passage (Wells et al.1996a).

Comparison to SAMOVA

We compare the performance of *BR* for the bottlenose dolphin dataset to the genetic clustering method *SAMOVA* (Dupanloup et al. 2002). Like *BR*, *SAMOVA* also incorporates a geographic constraint to ensure that the units defined by the method are geographically contiguous. *SAMOVA* also requires grouping samples into initial units, but uses a different measure of genetic differentiation than *BR* and incorporates simulated annealing to reduce the probability of finding a local rather than a global maximum. We ran *SAMOVA* using both the 'naïve' and 'expert' initial unit definitions used in the *BR* analyses. The geographic constraint used by *SAMOVA* is generated automatically. We used 10,000 simulated annealing steps and 10 different initial conditions. We repeated each analysis three times to ensure that we were obtaining consistent results.

Unlike *BR*, *SAMOVA* does not generated nested sets of hypothesized units. Rather, the user specifies the number of groups the method should define. If multiple hypotheses containing different numbers of units are desired, the analysis is run multiple times. We used *SAMOVA* to generate population structures containing 3, 4 and 5 groups. We then evaluated the performance

of the method by comparing the resulting structures to the observational and photo-identification studies.

Dispersal rate estimation

We estimate dispersal rates between several of the hypothesized units suggested by *BR* using *Migrate* (Beerli and Felsenstein, 2001). We first ran *Migrate* using the program default settings. The estimates of dispersal rate and ? resulting from this preliminary analysis were then used as the initial estimates in the final analysis. The final analysis again used the default search parameters, except that we used the 'randomtree = yes' option and replicated the analysis over three runs.

RESULTS

Boundary Rank **analyses**

BR analyses results (Figure 3) appear as bifurcating dendrograms with initial units at the tips. Nodes represent the coalescence of a pair of units with the genetic differentiation ($?^2$/dof) between the two units being coalesced shown above each node. Both naïve and expert analyses indicate that inland populations are more closely related to each other than they are to the animals from the Gulf of Mexico. Furthermore, in both analyses Sarasota Bay grouped with Charlotte Harbor before it did with Tampa Bay, suggesting greater gene flow between Sarasota Bay and Charlotte Harbor than between Sarasota Bay and Tampa Bay, despite the greater distance separating the former two bodies of water. These results match recent analyses of mtDNA and microsatellite data using an AMOVA approach and the same samples (Sellas et al. in prep.).

We used the program Migrate to estimate dispersal rates between Tampa Bay, Sarasota Bay, Charlotte Harbor and the Gulf of Mexico, as delineated by the *BR* analyses. The

delineation of the Sarasota Bay and Tampa Bay populations differed between the expert and naïve analyses, as shown in Figure 4. The dispersal rates between the inland water populations were all low, ranging from 2.2 to 7.7 dispersers per generation for the naïve analysis (Table 2) and from 0 to 8.4 in the expert analysis (Table 3). Dispersal estimates between the inland water populations and the Gulf population were also low for both analyses.

In the expert analysis, the Tampa Bay samples did not all group together; rather, the few samples from Boca Ciega Bay (Unit 1 in Figure 2b) remained separate from the samples from the rest of Tampa Bay until very late in the analysis. We therefore attempted to include Boca Ciega Bay as a separate stratum when estimating dispersal rates for the expert analysis, though the sample size from this area was low (n = 6). However, *Migrate* was unable to produce meaningful estimates of gene flow into or out of Boca Ciega Bay, with the 95% profile likelihood intervals ranging from zero to the implausibly high 10^8. Therefore, we excluded Boca Ciega Bay from the dispersal rate estimates.

SAMOVA **analyses**

The groupings suggested by *SAMOVA* were not consistent with the photo-identification and observational data (Figure 5). When the naïve initial units were used in the analysis, the samples from Sarasota Bay were all clustered into a distinct unit, but the remaining clusters did not correspond with actual populations. None of the clusters defined by the expert *SAMOVA* analysis corresponded with the observational data. Furthermore, *SAMOVA* was far more sensitive to the definition of initial units than was *BR*, as evinced by the fact that the *SAMOVA* results differed markedly depending on whether the naïve or expert initial units were used. When the expert initial units were used, the samples from Boca Ciega Bay remained distinct from the rest of Tampa Bay, as they did in the expert *BR* analysis.

DISCUSSION

The congruence between the results of the *BR* analyses and the pattern of population structure suggested by the observational data (Wells 1986, 1991, 2003; Scott et al. 1990, Wells and Scott 1990, Wells et al. 1996a, 1996b, 1997, Urian 2002) and previous genetic analyses (Duffield and Wells 1986, 1991, 2002; Sellas *et al.,* in prep.) indicate that *BR* is likely to prove a reliable tool for evaluating population structure for bottlenose dolphins in areas from which long-term observational and photo-identification data are not available. Differences between the units suggested by the naïve and expert analyses were mostly minor. Exact locations of the unit the Tampa Bay/Gulf of Mexico boundary and the Tampa Bay/Sarasota Bay boundary differed (Figure 5). The definition of initial units precluded identical placement in these cases. The magnitude and impact of boundary placement error depends on resulting biases in abundance estimation for the putative stocks. The Tampa Bay/Sarasota Bay boundary discrepancy is so small that abundance estimates error would be trivial. The Tampa Bay/Gulf of Mexico boundary differs more substantially between the analyses and warrants further investigation to actually estimate the magnitude of the error.

Of greater concern is the treatment of Boca Ciega Bay (Unit 1 from Figure 2), which was clustered with Tampa Bay in the naïve analysis, but remained distinct from the rest of Tampa Bay until very late in the clustering in the expert analysis. In fact, the expert analysis indicates that samples from Tampa Bay are more genetically distinct from Boca Ciega Bay than they are from Sarasota Bay and Charlotte Harbor (Figure 3). Figure 2 shows that in the naïve analysis, the six samples from Boca Ciega Bay were included in an initial unit containing eight samples from the deep-water portion of Tampa Bay. Thus, any genetic differentiation between Boca

Ciega Bay and the rest of Tampa Bay would have been obscured in the naïve analysis due to the configuration of the initial units.

The alleged strong differentiation between Boca Ciega Bay and Tampa Bay suggested by the expert analysis must be regarded with caution, however. The sample size from Boca Ciega Bay (n = 6) was the smallest of any of the initial units considered in either analysis (Table 1). Small sample size results in greatly increased uncertainty in the estimates of haplotype frequency. Thus, the fact that Boca Ciega Bay remained distinct from the rest of Tampa Bay until late in the expert analysis may simply be an artifact of its small sample size. Nevertheless, these results have prompted efforts to collect additional genetic samples in Tampa Bay, including Boca Ciega Bay, to further investigate the possibility of further population structure within Tampa Bay as suggested by the expert analysis in this paper and also by the behavioral work of Urian (2002).

Unlike *BR*, *SAMOVA* did not cluster the genetic samples into units consistent with observational data (Figure 6). Although *SAMOVA* has an apparent advantage over *BR* by using simulated annealing to avoid becoming trapped at a local minimum, *BR* uses as its measure of genetic differentiation a statistic that is more powerful for detecting population-level differences from mtDNA data (Hudson et al. 1992). This likely accounts for *BR*'s superior performance in correctly detecting population structure in the bottlenose dolphin dataset. In fact, the differentiation measure used by *SAMOVA*, F_{ct}, compares the proportion of variation among geographic regions to that between populations within the same region. Thus, *SAMOVA* is designed not for identifying populations, as it was used here, but rather for investigating the hierarchical relationships among populations. Consequently, the fact that it did not perform as well as *BR* in this application is likely a reflection of the fact that *BR* was specifically designed

for this type of problem while *SAMOVA* was not. *SAMOVA* may well outperform *BR* in addressing problems more comparable to that for which *SAMOVA* was designed.

The estimates of gene flow between Tampa Bay, Sarasota Bay, Charlotte Harbor and the coastal Gulf of Mexico were all demographically trivial, thus requiring that these areas be managed as separate management units. The reproductive lifespan of female bottlenose dolphins in Sarasota Bay is about 40 years (Wells and Scott 1999), with a generation of about 25-28 years, meaning that the per-generation movement rates in Tables 2 and 3 would all translate to less than a single animal dispersing per year. While such rates are sufficient to prevent the populations from maintaining independent evolutionary trajectories, they are so low as to have no meaningful impact on the demography of the populations. Thus, if one of these populations suffered a large natural or human-caused mortality event, the population would likely not be replenished by immigration from neighboring areas for several decades.

ACKNOWLEDGMENTS

Dolphin Quest, and NOAA Fisheries. Sampling and observational studies were performed under

Scientific Research Permits issued by NOAA Fisheries.

LITERATURE CITED

(* denotes references only cited in the 'Supplemental Information.')

Beerli, P. and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Science, USA. 98(8):4563-4568.

Duffield, D.A. and R.S. Wells. 1986. Population structure of bottlenose dolphins: Genetic studies of bottlenose dolphins along the central west coast of Florida. Contract Rept. to National Marine Fisheries Service, Southeast Fisheries Center. Contract No. 45-WCNF-5-00366. 10 pp.

Duffield, D.A. and R.S. Wells. 1991. The combined application of chromosome, protein and molecular data for the investigation of social unit structure and dynamics in *Tursiops truncatus*. Pp. 155-169 *In:* A.R. Hoelzel (ed.) Genetic Ecology of Whales and Dolphins. Report of the International Whaling Commission, Special Issue 13, Cambridge, U.K.

Duffield, D.A. and R.S. Wells. 2002. The molecular profile of a resident community of bottlenose dolphins, *Tursiops truncatus*. Pp. 3-11 *In*: C.J. Pfeiffer, ed., Molecular and Cell Biology of Marine Mammals. Krieger Publishing Company, Melbourne, FL.

Dupanloup, I., Schneider, S., Excoffier, L. (2002) A simulated annealing approach to define the genetic structure of populations. Molecular Ecology 11(12): 2571-81.

*Escorza-Trevino, S. and Dizon, A.E. 2000. Phylogeography, intraspecific structure, and sex-biased dispersal of Dall's porpoise, *Phocoenoides dalli*, revealed by mitochondrial and microsatellite DNA analysis. *Mol. Ecol.* 9:1049-60.

Excoffier, L., Smouse, P.E., and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479-91.

Fazioli, K. L. and R. S. Wells. 1999. Stock structure of coastal bottlenose dolphins, *Tursiops truncatus*, near Sarasota, Florida. Final Contract Report to National Marine Fisheries Service, Southeast Fisheries Science Center, Miami, FL. Contr. No. 40-WCNF701806

*Hudson, R.R., Boos, D.D. and Kaplan, N.L. 1992. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* 9(1):138-51.

*Latter, B.H.D. 1973. The island model of population differentiation: a general solution. *Genetics* 106:293-308.

*Lyrholm, T., Leimar, O. and Gyllensten, U. 1996. Low diversity and biased substitution patterns in the mitochondrial DNA control region of sperm whales: implications for estimates of time since common ancestry. *Mol. Biol. Evol.* 13:1318-26.

Martien, K.K and B.L. Taylor. 2003. Limitations…

Reynolds, J.E. III, R.S. Wells and S.D. Eide. 2000. Biology and conservation of the bottlenose dolphin. University Press of Florida.

Scott, M.D., R.S. Wells and A.B. Irvine. 1990. A long-term study of bottlenose dolphins on the west coast of Florida. Pp. 235-244 *In*: S. Leatherwood and R.R. Reeves (eds.), The Bottlenose Dolphin. Academic Press, San Diego. 653 pp.

Sellas, A.B. 2002. Population structure and group relatedness of bottlenose dolphins (*Tursiops truncatus*) in the coastal Gulf of Mexico using mitochondrial DNA and nuclear microsatellite markers. M. Sc. Thesis, University of California, Santa Cruz.

Sellas, A.B., R.S. Wells, and P.E. Rosel. In prep. Mitochondrial and nuclear DNA analyses reveal fine scale geographic structure in bottlenose dolphin (*Tursiops truncatus*) in the Gulf of Mexico.

*Takahata, N. and Palumbi, S.R. 1985. Extranuclear differentiation and gene flow in the finite island model. *Genetics* 109:441-57.

Taylor B.L. 1997. Defining "populations" to meet management objectives for marine mammals. pp. 347-364. *In*: A.E.Dizon, S.J.Chivers and W.F.Perrin (eds). *Molecular Genetics of Marine Mammals*. Society for Marine Mammalogy, Special Publication 3, Lawrence, Kansas.

*Taylor, B.L., Chivers, S.J., Sexton, S. and Dizon, A.E. 2000. Evaluating dispersal estimates using mtDNA data: comparing analytical and simulation approaches. *Conserv. Biol.* 14:1287-97.

*Taylor, B.L. and Dizon, A.E. 1999. First policy then science: why a management unit based solely on genetic criteria cannot work. *Mol. Ecol.* 8(Special Issue):S11-S16.

Urian, K.W. 2002. Community structure of bottlenose dolphins (*Tursiops truncatus*) in Tampa Bay, Florida, USA. M.Sc. thesis, University of North Carolina, Wilmington. 26 pp.

Wade, P.R. and R.P. Angliss. 1997. Guidelines for assessing marine mammal stocks: report of the GAMMS workshop, April 3-5, 1996, Seattle, Washington. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-OPR-12, 93 p.

Waring, G.T., D.L. Palka, P.J. Clapham, S.L. Swartz, M. Rossman, T. Cole, L.J. Hansen, K.D. Bisack, K. Mullin, R.S. Wells, D.K. Odell and N.B. Barros. 1999. U.S. Atlantic and Gulf of Mexico marine mammal stock assessments – 1999. NOAA Tech. Memo. NMFS-NE-153.

Wells, R.S. 1986. Population structure of bottlenose dolphins: Behavioral studies of bottlenose dolphins along the central west coast of Florida. Contract Rept. to National Marine Fisheries Service, Southeast Fisheries Center. Contr. No. 45-WCNF-5-00366.

Wells, R.S. 1991. The role of long-term study in understanding the social structure of a bottlenose dolphin community, p. 199-225. In: K. Pryor and K.S. Norris [eds.], Dolphin Societies: Discoveries and Puzzles. Univ. of California Press.

Wells, R.S. 1994. Determination of bottlenose dolphin stock discreteness: Application of a combined behavioral and genetic approach, p.16-20. In: K.R. Wang, P.M. Payne, and V.G. Thayer (compilers), Coastal Stock(s) of Atlantic Bottlenose Dolphin: Status Review and Management. Proceedings and Recommendations from a Workshop held in Beaufort, NC, 13-14 September 1993. NOAA Technical Memorandum NMFS-OPR-4.

Wells, R.S. 2003. Dolphin social complexity: Lessons from long-term study and life history. Pp. 32-56 *In:* F.B.M. de Waal and P.L. Tyack, eds., Animal Social Complexity: Intelligence, Culture, and Individualized Societies. Harvard University Press, Cambridge, MA.

Wells, R.S. and M.D. Scott. 1990. Estimating bottlenose dolphin population parameters from individual identification and capture-release techniques. Pp. 407-415 *In:* P.S. Hammond, S.A. Mizroch and G.P. Donovan (eds.) Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters. Report of the International Whaling Commission, Special Issue 12, Cambridge, U.K. 440 pp.

Wells, R.S. and M.D. Scott. 1999. Bottlenose dolphin *Tursiops truncatus* (Montagu, 1821). Pp. 137-182 *In:* S.H. Ridgway and R. Harrison (eds.), Handbook of Marine Mammals, Vol. 6, the Second Book of Dolphins and Porpoises. Academic Press, San Diego, CA. 486 pp.

Wells, R.S. and M.D. Scott. 2002. Bottlenose dolphins (*Tursiops truncatus* and *T. aduncus*). Pp. 122-128 *In:* W.F. Perrin, B. Würsig, and J.G.M. Thewissen, eds., Encyclopedia of Marine Mammals. Academic Press, San Diego, CA.

Wells, R.S., M.D. Scott and A.B. Irvine. 1987. The social structure of free-ranging bottlenose dolphins. Pp. 247-305 *In*: Genoways, H. (ed.), Current Mammalogy, Vol. 1. New York: Plenum Press.

Wells, R.S., M.K. Bassos, K.W. Urian, W.J. Carr and M.D. Scott. 1996a. Low-level monitoring of bottlenose dolphins, *Tursiops truncatus*, in Charlotte Harbor, Florida: 1990-1994. NOAA Tech. Mem. NMFS-SEFSC-384.

Wells, R.S., M.K. Bassos, K.W. Urian, S.H. Shane, E.C.G. Owen, C.F. Weiss, W.J. Carr and M.D. Scott. 1997. Low-level monitoring of bottlenose dolphins, *Tursiops truncatus*, in Pine Island Sound, Florida: 1996. Final Contract Report to National Marine Fisheries Service, Southeast Fisheries Science Center, Miami, FL. Contr. No. 40-WCNF601958.

Wells, R.S., H.L. Rhinehart, L.J. Hansen, J.C. Sweeney, F.I. Townsend, R. Stone, D. Casper, M.D. Scott, A.A. Hohn, and T.K. Rowles. In press. Bottlenose dolphins as marine ecosystem sentinels: Developing a health monitoring system. Ecohealth 1 (2).

Wells, R.S., K.W. Urian, A.J. Read, M.K. Bassos, W.J. Carr and M.D. Scott. 1996b. Low-level monitoring of bottlenose dolphins, *Tursiops truncatus*, in Tampa Bay, Florida: 1988-1993. NOAA Tech. Mem. NMFS-SEFSC-385.

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. Sixth Int. Congr. Genet.* 1:356-66.

York A.E., Merrick R.L. and Loughlin T.L. 1996. An analysis of the Steller sea lion metapopulation in Alaska. pp. 259-292. *In*: D.R.McCullough (ed.). *Metapopulations and Wildlife Conservation*. University of California Press.

LIST OF FIGURES

Figure 1. Genetic samples were collected from dolphins from Tampa Bay (*n*=46; red), Sarasota Bay (*n*=54; yellow), Charlotte Harbor (*n*=51; green) and nearby coastal Gulf of Mexico waters (*n*=56; blue) (Sellas *et al.*, in prep.).

Figure 2a&b. The initial units used in the (a) naïve and (b) expert analyses. Connectivity is indicated by the red lines.

Figure 3. Bifurcating dendrograms depicting the results of the *Boundary Rank* analyses using the (a) naïve and (b) expert initial unit definitions. The initial units are at the tips of the dendrograms. The nodes represent the coalescence of a pair of units, while the number associated with each node indicates the genetic distance ($?^2$/dof) between the two units being coalesced. The depth of a node within the dendrogram indicates how late in the analysis the two units connected to that node coalesced. For instance, in the naïve analysis, the first pair of units to combine was two initial units from Sarasota Bay, numbers 7 and 9. Next units 8 and 11, both from the Gulf of Mexico, coalesced. The number of units remaining at any level in the dendrogram is indicated along the right edge of the figure.

Figure 4. The putative populations between which dispersal was estimated using the program *Migrate*. For the (a) naïve analysis, all Tampa Bay samples were treated as a single population, while for the (b) expert analysis Boca Ciega Bay was treated separately from the rest of Tampa Bay.

Figure 5. Results of the *SAMOVA* analyses of the bottlenose dolphin dataset. *SAMOVA* was used to group the samples into 3, 4 or 5 units using both the 'naïve' and 'expert' initial units.

Table 1. Sample sizes used for the (a) naïve and (b) expert analyses.

Initial Unit	Sample Size
a) Naïve analysis	
1	14
2	9
3	9
4	17
5	7
6	14
7	20
8	18
9	17
10	12
11	19
12	22
13	29
14	33
b) Expert analysis	
1	6
2	37
3	9
4	22
5	13
6	20
7	11
8	8
9	13
10	13
11	38
12	30

Table 2. Pairwise estimates from *Migrate* of gene flow between the three inland-water strata and the coastal Gulf of Mexico as defined by the 'naïve' *BR* analysis. All estimates are expressed in terms of the number of effective female dispersers per generation, with the 95% profile likelihood interval given in parentheses. Note that rates of gene flow are not symmetric. TB = Tampa Bay, SB = Sarasota Bay, CH = Charlotte Harbor and GM = Gulf of Mexico.

Target population	Source population			
	TB	SB	CH	GM
TB	--	4.6	4.6	6.4
		(2.5-7.6)	(2.5-7.6)	(3.9-9.8)
SB	6.7	--	7.2	4.1
	(3.7-11.0)		(3.1-11.7)	(1.9-7.7)
CH	7.7	2.2	--	2.6
	(4.9-11.5)	(0.9-4.5)		(1.1-5.0)
GM	8.4	2.8	5.2	--
	(5.3-12.6)	(1.2-5.4)	(2.9-8.6)	

Table 3. Pairwise estimates from *Migrate* of gene flow between the three inland-water strata and the coastal Gulf of Mexico defined by the 'expert' *BR* analysis. All estimates are expressed in terms of the number of effective female dispersers per generation, with the 95% profile likelihood interval given in parentheses. Note that rates of gene flow are not symmetric. TB = Tampa Bay, SB = Sarasota Bay, CH = Charlotte Harbor and GM=Gulf of Mexico.

Target population	Source population			
	TB	SB	CH	GM
TB	--	1.6	8.8	4.0
		(0.8-2.9)	(6.5-11.4)	(2.6-5.9)
SB	0.6	--	2.4	0
	(0.2-1.4)		(1.5-3.6)	(0-0)
CH	7.9	1.2	--	0.7
	(6.0-10.1)	(0.6-2.1)		(0.2-1.5)
GM	4.8	1.9	6.9	--
	(2.8-7.5)	(0.7-3.9)	(4.3-10.3)	

Supplemental Information

SIMULATION PERFORMANCE TESTING OF *BOUNDARY RANK*

Performance testing framework

We used data generated by the simulation model described in Taylor *et al*. (2000) to conduct a systematic evaluation of the performance of *Boundary Rank (BR)*. We used the simulation model to generate datasets for which we knew the true population structure. We drew samples from each of the model populations, divided those samples into discrete sampling sites, and then analyzed the resulting samples using the *Boundary Rank* method. All of our simulations consisted of five populations arranged as a linear stepping-stone, with dispersal occurring only between adjacent populations. We used a connectivity matrix in which each sampling site was 'connected' only to its two adjacent neighbors. We evaluated the performance by comparing the locations of the four highest-ranking boundaries from the analysis to the locations of the four actual population boundaries between the five model populations. Thus, our performance measures examine only the ranking of the boundaries, not the likelihood of the researcher choosing the correct number of boundaries to include.

The Simulation Model

The model used to generate the data for evaluating the performance of *BR* is described in Taylor *et al*. (2000). It is a density-dependent birth-death model that simulates the evolution of mitochondrial DNA. We simulated 40 variable basepairs of mitochondrial sequence and used birth and death rates that resulted in a generation time of approxima tely four years. The mutation parameter was set at 0.0001 per basepair per generation. This value was chosen to produce haplotype distributions that match those we observe for northern right whale dolphins (*Lissodelphis borealis*), a species for which we have a large amount of sequence data.

For each simulation, all five populations were initialized at carrying capacity with a single haplotype and allowed to evolve for 200,000 years. After 150,000 years, stochastic equilibrium was reached. The haplotype profile (the sequence of each haplotype and its frequency in all five populations) was recorded every 500 years between years 150,000 and 200,000 of the simulation, providing 100 haplotype profiles for each combination of dispersal rates and carrying capacities. After a haplotype profile had been generated, samples were drawn from each of the five populations and divided into multiple sampling sites. This sampling was repeated five times for each of the 100 haplotype profiles, resulting in 500 independent sampling events for each combination of dispersal rate and carrying capacity. For each sampling event, *Boundary Rank* was used to rank all putative boundaries. The locations of the four highest-ranking boundaries were then compared to the locations of the four actual population boundaries.

Measuring Performance

We measured the performance of the method in two ways. First, we determined the probability that the four highest-ranking boundaries from the analysis exactly correspond to the actual population boundaries. We call this the "exact placement measure." The goal of most attempts to define conservation units, however, is not to get the unit boundaries exactly right, but rather to meet some predefined management objective. We therefore decided to use a second measure of performance to determine how well our method would perform under a specific real-world management scheme. We call this the "management scheme measure." We chose the management scheme applied to management units for marine mammals in the United States under the Marine Mammal Protection Act (MMPA). One of the objectives of the MMPA is that a species must remain extant throughout its entire range with no local extirpations (Taylor, 1997). Human-caused mortality (either direct harvest or incidental mortality) is allowed for

marine mammal species so long as it does not exceed the allowable kill (Taylor, 1997). The calculation of allowable kill uses the estimated abundance of each management unit and is therefore vulnerable to errors when units are improperly defined.

The consequences of over-estimating the size of the exploited population will depend on the spatial distribution of human impact. The more evenly distributed the impact is across a management unit, the less severe the consequences of defining management units that are larger than the actual populations (Taylor and Dizon, 1999). We therefore developed a stringent test where human-caused mortality was assumed to be maximally concentrated to a single point location (e.g., by-catch in a highly concentrated fishery).

The four highest-ranking boundaries from the *BR* analysis divided the samples into five hypothesized management units. To simulate the consequences of managing on the basis of these five units we used a deterministic model of five populations exhibiting logistic growth. Dispersal occurred between adjacent populations at rate *d*, and harvest was removed only from the focal sampling site. We assumed that harvest occurred at the maximum rate allowed under the Potential Biological Removal (PBR; Taylor, 1997) scheme, which is the management procedure used to regulate human-caused mortality under the MMPA. The abundance estimate that we used to calculate the allowable harvest was the abundance of the management unit containing the focal sampling site. Thus, if the management unit were larger than the actual population then the abundance would be overestimated. If this overestimate were substantial, the harvested population would be driven to extinction.

We ran the deterministic model for 10,000 years (by which time the populations had reach equilibrium abundance) to see whether the management unit boundaries suggested by *BR* were sufficiently close to the true population boundaries to prevent the extirpation of the

harvested population. We repeated this process for each sampling site and recorded whether or not harvest from each sampling site would result in extinction. We calculated performance for each population by averaging the results for all of the sampling sites contained within a population. In the example (Fig. S1), the performance for populations 1, 2 and 3 is 1.0, since harvesting from any of the sampling sites within those populations would not cause a local extinction. Performance for population 5 is 0.5, since two of the four sampling sites from population 5 are contained in a management unit that is 50% too large and is therefore not sufficient to prevent the over-exploitation and eventual extinction of the harvested population. Similarly, since it is entirely within a management unit too large to adequately protect it, performance for population 4 is zero. The overall performance, averaged across populations, is 0.7.

By examining the impact of harvest at every sampling site and averaging the results, our performance measure takes account of the uncertainty in the sampling location. This performance measure is also conservative since we assume that not only is the human-caused mortality concentrated at a single location but also that harvest occurs at its maximum allowable rate.

Equal abundance and dispersal rate and even sampling

We first tested performance assuming that each of the five model populations had the same abundance and that the dispersal rate was equal between all pairs of adjacent populations. The expected level of genetic differentiation between two populations is given by the formula

$$F_{ST} = \frac{1}{2NdT + 1} \tag{1}$$

where N is the effective abundance of the populations (for mitochondrial data, roughly equal to the number of adult females), d is the annual dispersal rate and T is the generation time (Wright,

1932; Latter, 1973; Takahata and Palumbi, 1985). We express gene flow in terms of an annual dispersal rate rather than the more familiar migrants-per-generation (*Nm*) because the annual dispersal rate is the measure relevant to management. As discussed by Taylor (1997) and Taylor and Dizon (1999), dispersal rate is the critical parameter in determining whether two areas can be safely managed as a single unit. To determine the impact of the degree of genetic differentiation on the performance of *Boundary Rank*, we examined abundances of 100, 300, and 1,000 effective adult females and dispersal rates ranging from 0 to 0.01 per year. We also examined performance as a function of the total number of samples (120, 180 or 240) and the total number of sampling sites (10, 15 or 20).

The performance of the method, as measured by both the exact placement and management scheme measures, was consistently highest when we used either $?^2$/dof as our measure of genetic similarity rather than F_{ST} or ϕ_{ST} (Fig. S2). This result is consistent with published studies that have shown that $?^2$ is the most powerful statistic for detecting population structure using mtDNA data (Hudson *et al.*, 1992). The superior performance of the $?^2$-based measure was consistent across all of the abundances, dispersal rates, and sample distributions we examined. All results presented in the remainder of this paper were obtained using $?^2$/dof as the similarity measure.

Performance using the exact placement measure was high when dispersal rates were low, but declined with increasing dispersal rate (Fig. S3a-c). Performance decreased substantially with fewer samples (Fig. S3a) and with more sampling sites (Fig. S3b), both of which result in fewer samples per sampling site. Performance also decreased with increasing abundance, though the difference in performance between abundances of 300 and 1,000 was much greater than the decrease between 100 and 300 (Fig. S3c).

Boundary Rank performed much better when evaluated using the management scheme measure, which strives to keep all local populations extant even in a worst case where all of the human-caused mortality is concentrated at a single point. Performance decreased slightly with increasing annual dispersal rate up to a rate of between 0.006 and 0.008, after which it increased slightly (Fig. S3d-f). Overall, however, the decrease in performance with increasing dispersal was much less pronounced than with the exact placement measure (comparing the figures on the left to their equivalents on the right in Fig. S3). Decreasing the number of samples resulted in a slight decrease in performance (Fig. S3d), as did increasing the number of sampling sites (Fig. S3e). As expected, an increase in the abundance of the populations caused performance to decrease more rapidly with increasing dispersal rate (Fig. S3f) due to the inverse relationship between abundance and genetic differentiation (eq. 1). However, none of these changes had a dramatic impact on performance.

The difference in performance between the two measures is due to differences in the way that errors in boundary placement are weighted. The exact placement measure (getting boundaries exactly right) weights all errors in boundary placement equally and consequently ignores the magnitude of the error. However, when actually defining units of conservation, some errors in unit definition will be more serious than others. A small error, one in which a boundary is placed very near to an actual population boundary, is likely to have little effect on the outcome of management decisions. Large errors in boundary placement, on the other hand, could result in serious management errors. The management scheme measure takes into account the severity of boundary placement errors. As long as an error in boundary placement is small enough to meet management objectives, the error does not translate into reduced performance under the second measure. Thus, the disparity between the two measures at high dispersal rates indicates that

while boundaries placed by the analysis do not always exactly correspond to actual population boundaries, they are usually very close. Because the management scheme measure is more relevant to performance in a management context, we focus on this measure in subsequent performance testing.

The increase in performance using the management scheme measure when the annual dispersal rate increased from 0.008 to 0.01 is due to the nature of the management scheme we chose and its management objectives. Because the goal is simply to prevent the extinction of local populations, the cost of making an error in boundary placement goes down as dispersal rate increases. This is because, when the dispersal rate between two populations is high, over-exploitation in one population can be made up for by dispersal from the other population and larger errors in boundary placement can be accommodated without resulting in decreased performance. Thus, while the accuracy of boundary placement decreases with increasing dispersal rate, as evinced by the exact placement measure, the decrease in accuracy is partially mitigated by a concomitant decrease in the cost of boundary placement errors. In this case, once the annual dispersal rate gets over approximately 0.008, the decrease in cost outweighs the decrease in accuracy, and performance actually increases.

Unequal abundance and dispersal

In reality, natural populations are not all the same size, nor do they share dispersers at an equal rate. We therefore looked at the effect of unequal abundance and dispersal rate on performance using the abundances and dispersal rates shown in Figure S4. Initially we tested performance for the populations shown in Figure S4 assuming that sampling effort was evenly distributed across the range. For this analysis we drew a total of 250 samples and divided them among 25 sampling sites (10 samples per site). The number of sampling sites from each

population was proportional to the effective abundance of the population (Table S1). We found that the probability that the units defined by the four highest-ranking boundaries would adequately protect populations 2 through 5 ranged from 0.766 to 0.998 (Table S1). However, performance for population 1 was only 0.384, meaning that it would be protected from over-exploitation only 38.4% of the time if it were harvested under the PBR management scheme. This is because population 1 is a small population ($N = 100$) situated next to a population ten times larger with which it exchanges dispersers at a rate of 0.005 per year. Thus, while population 1 only exports an average of 0.5 (100 x 0.005) dispersers per year, it receives 5 (1,000 x 0.005) dispersers per year from population 2. Consequently, genetic drift within population 1 is swamped by immigration from population 2, preventing any detectable genetic differentiation from developing. Thus, neither *Boundary Rank* nor any other genetic approach will have much success in distinguishing populations 1 and 2.

Fortunately, the high rate of immigration into population 1 means that the consequences of failing to detect the boundary that defines it are less severe than if the population received fewer migrants. Because gene flow is so high into this small population, no local adaptation would be expected without strong selection. Thus, the extirpation of population 1 would result in little if any loss of genetic diversity. Furthermore, if the exploitation of population 1 were halted after the population had been extirpated, individuals from population 2 would quickly recolonize the region. Thus, the severity of the error in underestimating population structure in this area is less than if the boundary represented low gene flow and may be an acceptable temporary loss depending on other factors, such as whether the population was of cultural or economic importance (such as use for ecotourism).

In contrast, although loss of the other small, isolated population (population 5) is less likely (protected 76.6% of the time compared to 38.4% of the time for population 1), the consequences are more severe because gene flow is sufficiently low that local adaptation is plausible and recolonization would be very slow. This example emphasizes that even when dispersal is very low (only two dispersers per generation move from population 4 to population 5) and a relatively large proportion (10%) of the populations is sampled, it is still extremely difficult to detect a small population situated next to a much larger one.

Uneven sampling

Because we wanted a realistic assessment of how our method would perform when faced with real-world data, we also tested performance using two different examples of unevenly distributed samples: (1) the "concentrated" distribution had a high density of sampling sites near the center of the range and low density near the edges (Fig. S5a) and (2) the "gappy" distribution had high sample site densities near the edges and a low density sampling gap near the center (Fig. S5b). We first tested performance with each of these uneven sampling distributions when all five model populations had the same abundance ($N = 1,000$) and exchanged dispersers with their neighbors at an equal rate ($d = 0.004$). This constituted a rigorous test, as that combination of abundance and dispersal resulted in one of the lowest estimates of performance when sampling sites are evenly distributed (Table S1). Second, we tested performance using each of the two uneven sampling distributions (concentrated and gappy) and the abundances and dispersal rates shown in Figure 4. This is our most rigorous test of performance, since it includes unequal abundance and dispersal rates and uneven sample distribution.

For both equal (Table S1) and unequal (Table S2) population abundances and dispersal rates, performance was largely a function of the number of sampling sites (and therefore,

samples) per population. When the number of sampling sites from a population was three or greater, performance was generally over 0.75. When the number of sampling sites was less than three, performance was generally less than 0.6. This result illustrates the importance of careful sample collection in an investigation of population structure. When samples are unevenly distributed across the study area, as they often are, it is unlikely that population structure will be detected in areas where the sampling is poorest. Consequently, sampling efforts should be concentrated in areas where human impact poses the greatest threat to a species.

Summary

While the probability of placing boundaries in exactly the right location, as measured by the exact placement performance measure, decreases substantially with increasing dispersal rate, the management scheme performance measure shows that the four highest-ranking boundaries from the analysis are usually close enough to the actual population boundaries to adequately protect against over-exploitation due to errors in unit definition.

It is worth noting that all of the simulated datasets used in the performance testing were designed to have haplotypic diversities comparable to those found in northern right whale dolphins (*Lissodelphis borealis*). However, haplotypic diversities vary widely among species of cetaceans, from very low levels in sperm whales (*Physeter macrocephalus*; Lyrholm *et al.*, 1996) to very high levels in species such as Dall's porpoise (*Phocoenoides dalli*; Escorza-Trevino and Dizon, 2000). Because the *Boundary Rank* method relies on haplotypic frequency data in calculating the genetic distance between units, the sample size necessary to achieve a given level of performance will increase as haplotypic diversity increases. Thus, when evaluating the adequacy of a given sample for use in *Boundary Rank* or any other analysis program consideration must be given to not only the number of samples but also the sample size relative

to haplotypic diversity. For instance, while 20 samples per sampling site may be judged sufficient if only five to ten haplotypes are detected at each sampling site, such a sample size will not yield sufficiently accurate frequency information if most sampling sites contain 15 to 20 haplotypes, most of which are represented by a single individual.

FIGURE LEGENDS

Figure S1. Schematic of the procedure used to estimate performance under the management scheme measure. In the sample shown, 30% of the range is contained in a management unit that is too large to prevent over-exploitation, so the performance is 0.7.

Figure S2. Performance of *Boundary Rank* as a function of dispersal rate using different measures of genetic similarity. The results shown were generated by drawing a total of 240 samples from the 5 model populations and dividing the samples among ten sampling sites. The abundance of each of the five model populations was 300 effective adult females. Performance was measured using both the a) exact placement and b) management scheme measures.

Figure S3. Performance of the analysis as a function of dispersal rate with equal abundances and dispersal rates and an even sample distribution. Panels a through c show the results using the "exact placement" measure of performance and panels d through f show the results for the "management scheme" measure of performance. We started with a total of 240 samples divided among ten sampling sites and the abundance of each of the five populations equal to 300 effective adult females. We then independently varied the total number of samples (a and d), the number of sampling sites (b and e), and the effective abundance of the populations (c and f).

Figure S4. Effective abundances and dispersal rates used when testing the performance of the method in the face of unequal abundance and dispersal rates.

Figure S5. Two examples of unevenly distributed sampling sites: a) a "concentrated" sampling distribution, where most sampling sites are concentrated near the center of the range, and b) a "gappy" sampling distribution, where most sampling sites are near the edges of the range with low sampling site density near the center. Both distributions consisted of 20 sampling sites with 12 samples per site, for a total of 240 samples.

Table S1. Performance of the ranking method, using the management scheme measure, as a function of sample distribution when all populations have an abundance of 1,000 and exchange dispersers at a rate of 0.004 per year. The results shown are based on 20 sampling sites, each comprised of 12 samples. For each sampling scheme we list the performance as measured by the management scheme measure, the number of sampling sites from each population and the percentage of the total abundance of each population that was sampled.

Sample Distribution		Population (N_e)					Average
		1 (*1,000*)	2 (*1,000*)	3 (*1,000*)	4 (*1,000*)	5 (*1,000*)	
Even	Performance	0.842	0.741	0.734	0.776	0.850	0.788
	# of sites	4	4	4	4	4	
	Proportion sampled	0.048	0.048	0.048	0.048	0.048	
Concentrated	Performance	0.929	0.965	0.845	0.402	0.360	0.700
	# of sites	3	9	5	2	1	
	Proportion sampled	0.036	0.108	0.06	0.024	0.012	
Gappy	Performance	0.842	0.478	0.318	0.632	0.896	0.633
	# of sites	8	2	1	3	6	
	Proportion sampled	0.096	0.024	0.012	0.036	0.072	

Table S2. Performance of the ranking method, using the management scheme measure, as a function of sample distribution when population sizes and dispersal rates are unequal. For the even sample distribution, there were 25 sampling sites, each containing 10 samples. For the concentrated and gappy distributions, there were 20 sampling sites of 12 samples each. For each sampling scheme we list the performance as measured by the management scheme measure, the number of sampling sites from each population and the proportion of the total abundance of each population that was sampled.

Sample Distribution		Population (N_e)					Average
		1 (*100*)	2 (*1,000*)	3 (*300*)	4 (*1,000*)	5 (*100*)	
Even	Performance	0.384	0.998	0.805	0.884	0.766	0.767
	# of sites	1	10	3	10	1	
	Proportion sampled	0.1	0.1	0.1	0.1	0.1	
Concentrated	Performance	0.496	1.0	0.756	0.824	0.830	0.781
	# of sites	1	10	3	5	1	
	Proportion sampled	0.1	0.1	0.1	0.05	0.1	
Gappy	Performance	0.56	0.998	0.528	0.652	0.947	0.737
	# of sites	2	8	1	6	3	
	Proportion sampled	0.2	0.08	0.03	0.06	0.33	

Figure 1.

Tampa Bay

Sarasota Bay

Gulf
of
Mexico

Charlotte
Harbor

Figure 2a.

Naïve Analysis
Initial Units

Gulf of
Mexico
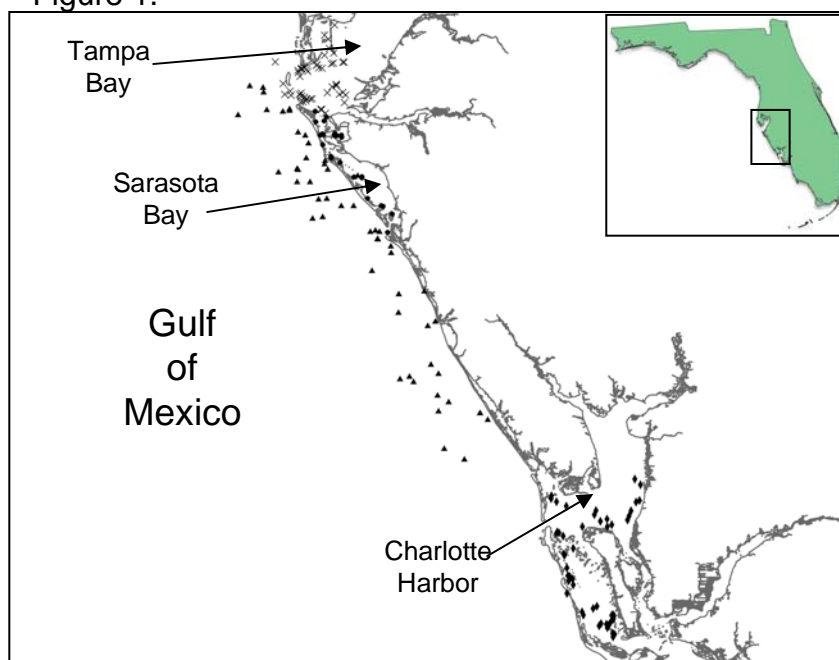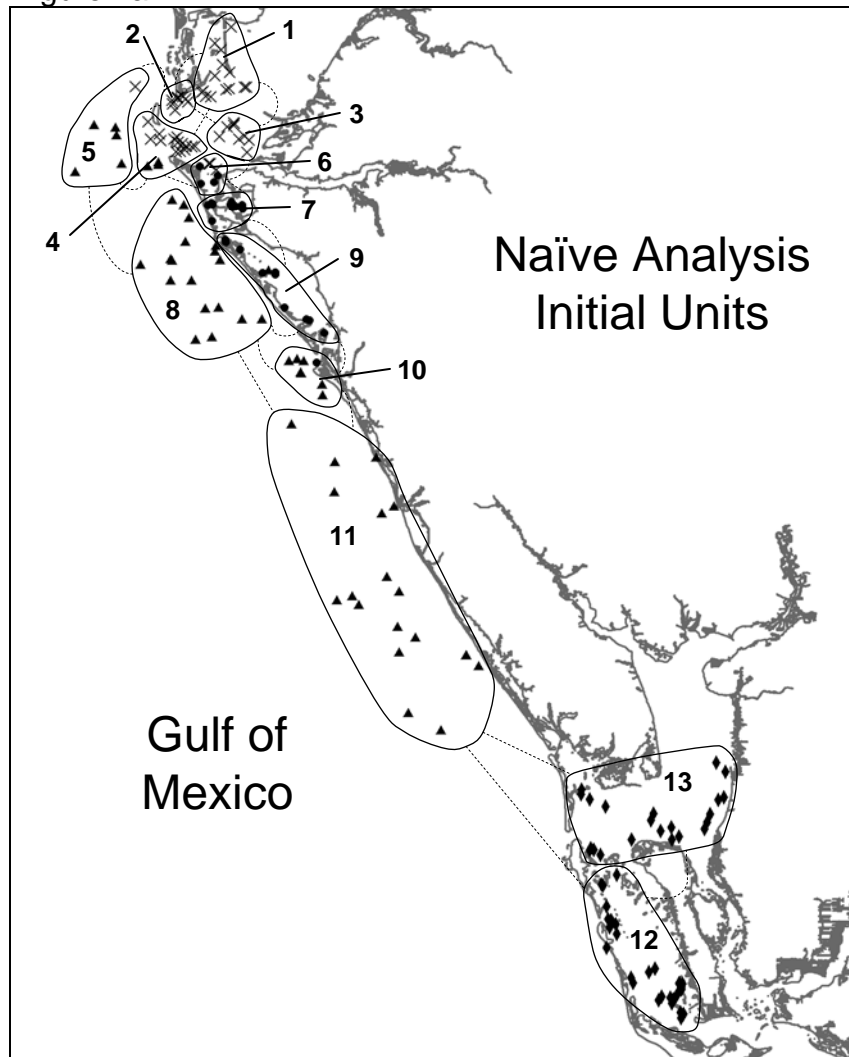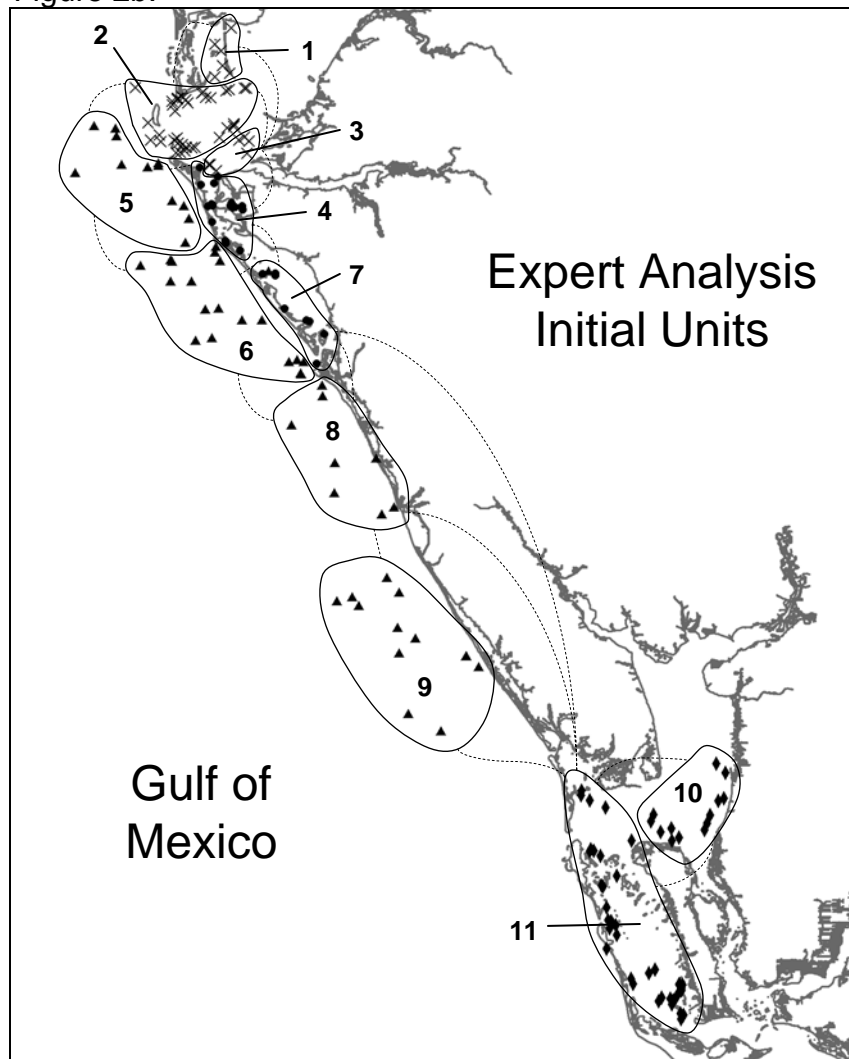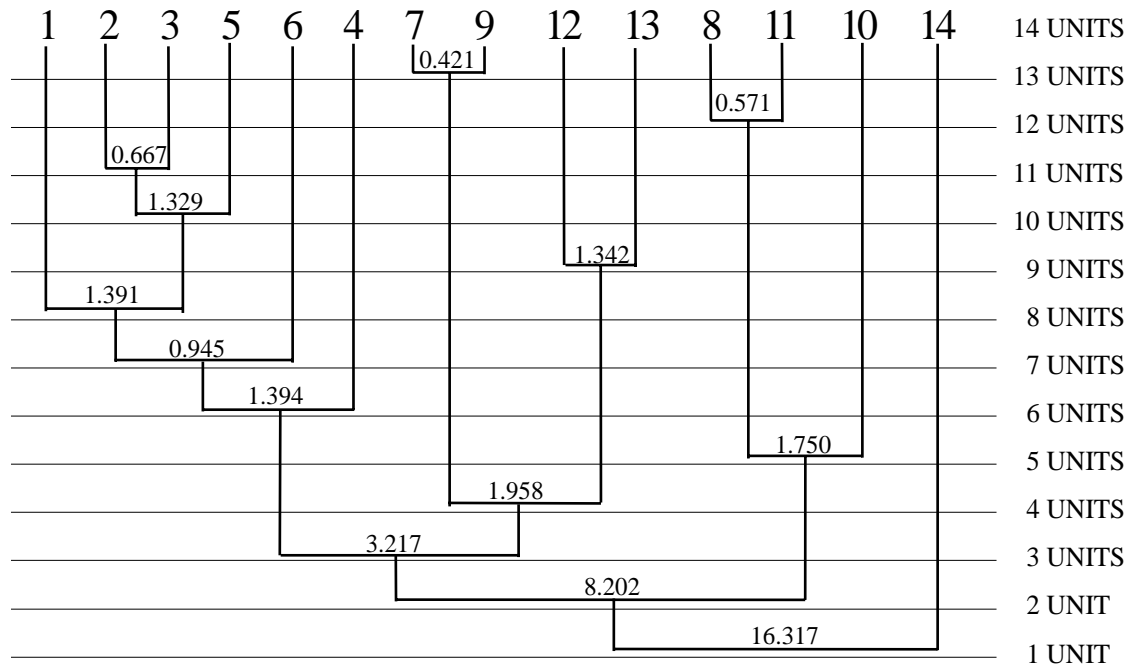
Figure 2b.
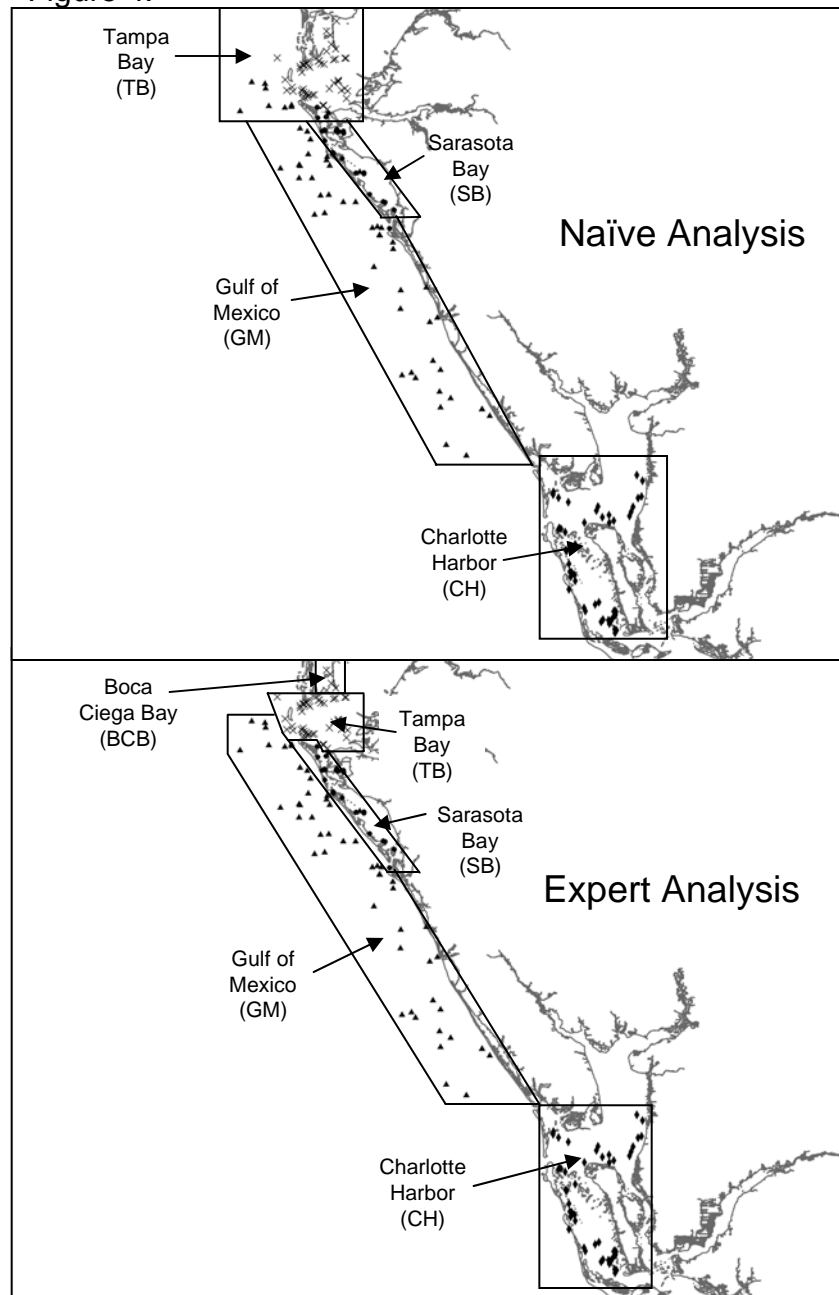
Figure 3.

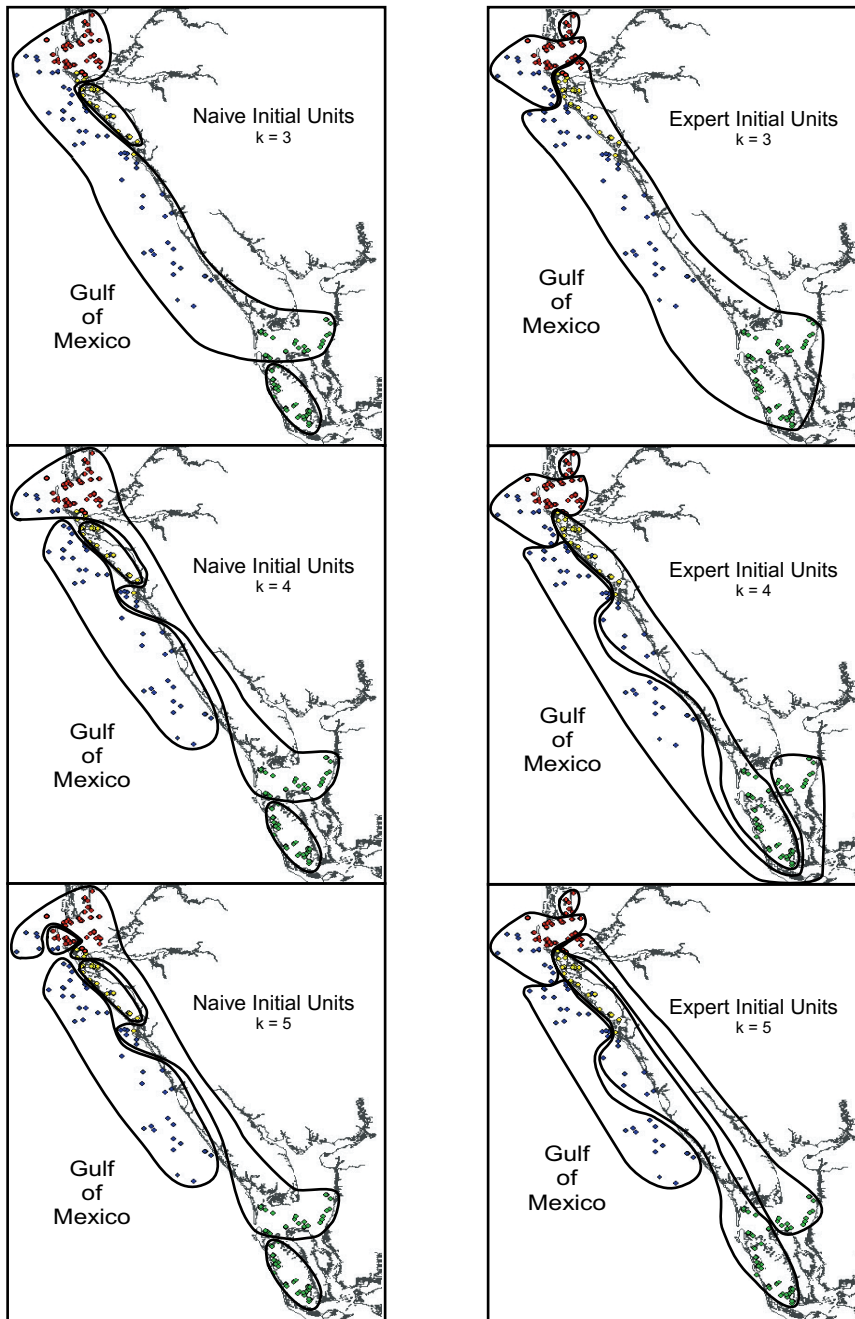## Naive analysis

Expert analysis
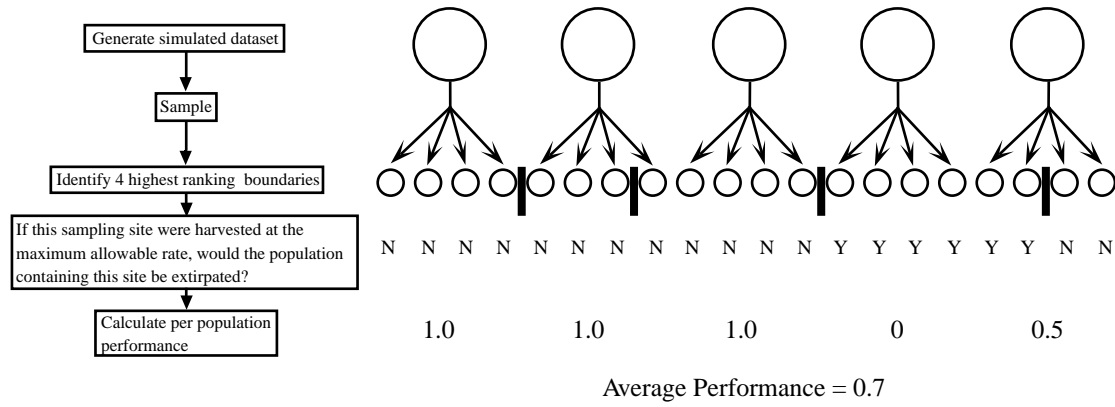
Figure 4.

Figure 5.

Naive Initial Units
k = 3

Gulf
of
Mexico

Expert Initial Units
k = 3

Gulf
of
Mexico

Naive Initial Units
k = 4

Gulf
of
Mexico

Expert Initial Units
k = 4

Gulf
of
Mexico

Naive Initial Units
k = 5

Gulf
of
Mexico

Expert Initial Units
k = 5

Gulf
of
Mexico

Figure S1.

Generate simulated dataset

Sample

Identify 4 highest ranking boundaries

If this sampling site were harvested at the maximum allowable rate, would the population containing this site be extirpated?

Calculate per population performance

N N N N N N N N N N N N Y Y Y Y Y Y N N

1.0   1.0   1.0   0   0.5

Average Performance = 0.7

Figure S2.

a)

Exact Placement

b)

Management Scheme

Figure S3.

Figure S4.

	High		Low	High		Low	
$d =$	0.005		0.0005	0.005		0.0005	

	1	2	3	4	5

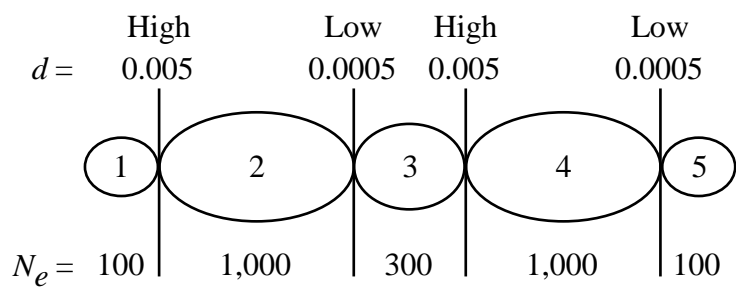| $N_e =$ | 100 | 1,000 | 300 | 1,000 | 100 |

Figure S5.

a) Concentrated

●   ●   ● ●●●◉ ●●● ● ●●   ●●   ●   ●   ●   ●

b) Gappy

● ●●●●● ● ●●   ●   ●   ●   ● ●   ●● ● ●● ●